

September 2016

Locally Developed Performance Assessments: One State's Decision to Supplant Standardized Tests with Alternative Measures

Amy Lynn Abbott

Old Dominion University, alabbott77@gmail.com

Follow this and additional works at: <http://digitalcommons.gardner-webb.edu/joel>

 Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Educational Methods Commons](#)

Recommended Citation

Abbott, Amy Lynn (2016) "Locally Developed Performance Assessments: One State's Decision to Supplant Standardized Tests with Alternative Measures," *Journal of Organizational & Educational Leadership*: Vol. 2: Iss. 1, Article 5.
Available at: <http://digitalcommons.gardner-webb.edu/joel/vol2/iss1/5>

Locally Developed Performance Assessments: One State's Decision to Supplant Standardized Tests with Alternative Measures

Cover Page Footnote

About the Author: Amy L. Abbott received her Ph.D. from the Department of Teaching and Learning at Old Dominion University in Norfolk, Virginia. She has served as an instructor at the university teaching undergraduate courses to pre-service teachers, as well as a district professional development trainer with topics such as performance tasks and rubric development, differentiation, and data-informed decision making. Her research interests include implementation of alternative assessment formats as they relate to policy reform, alongside the role of formative assessment in classroom practice and instructional planning.

Acknowledgments: I would like to acknowledge Dr. Brandon Butler, Dr. Jamie Colwell, and Dr. Steve Myran for their expertise during the development of this study and manuscript.

INTRODUCTION

School districts in the United States are increasingly considering alternative assessments to assess student achievement and ensure students are well equipped for workforce demands and college preparedness. A prime example of this was the 2014 Virginia General Assembly legislation in the removal of five, end-of-year Virginia Standards of Learning tests from select elementary and middle school grades/subjects and supplant them with alternative measures (House Bill 930/Senate Bill 306). As a result, local districts were encouraged to capitalize on alternative formats of assessment that would provide students with innovative opportunities to demonstrate proficiency with concepts, skills, and content knowledge (Virginia Department of Education-VDOE, 2014).

According to the legislative guidelines outlined in the State Superintendent's Memo (October 24, 2014), three elementary school SOL tests – Grade 3 History, Science, and Grade 5 Writing – and two middle school SOL tests – Grade 6 US History to 1865 and Grade 7 US History: 1865 to the Present – were removed. Notably, the legislation granted each district's school board the discretion to decide on the alternative assessment format, to include choices of performance and project-based assessments with rubrics, portfolios, or other integrated tests (VDOE, 2014). These alternative assessment designs were intended to allow for assessment of both process and product (Basturk, 2005; Moon, Brighton, Callahan, & Robinson, 2005), while requiring students to apply knowledge and provide rationale for their thinking through open-response formats (Khattri, Reeve, & Kane, 1998). Such assessments are perceived to be authentic “when they are modeled after real-life problems and tasks and can supplement or replace conventional paper-and-pencil test” (Martin-Kniep, Sussman, & Meltzer, 1995, p. 47). When appropriate, the legislation encouraged the development of interdisciplinary assessments (i.e., social studies and English) that incorporate real-life situations and contexts.

As a result of the removed standardized tests, Virginia districts were tasked with providing evidence as to whether or not students' proficiency of content knowledge and

cognitive skills could be attained through alternative means. As outlined by the legislative guidelines, the expanded use of authentic assessments signified a new direction for Virginia. First, the need for ongoing professional development during the early years of the reform was encouraged by the state to promote high-quality instruction geared toward students' preparation and completion of the assessments (VDOE, 2014). Moreover, the legislation called for an evaluation of teachers' readiness and capacity to design and implement the alternative measures, to which the results were used in the structuring of professional development around associated needs for support. Next, the state intended for the reform to encourage collaboration of district leaders, administrators, coordinators, and teachers, across schools and districts throughout Virginia. Lastly, local school boards and superintendents were to play a vital role in ensuring compliance of the state's mandates through the submission of an annual Standards of Quality report outlining the district's plan of action.

Significant to the legislation's enactment was the concern for the invested time spent preparing students for standardized tests. Since the inception of the state Standards of Learning (SOL) tests in 1998, teachers have adapted instruction and assessment based upon the increased emphasis on standardized testing, starting in third grade. Resulting from the thrust for a national increase in standardized testing in the 1990s, research reflects questions of its effectiveness upon considering the pressures placed upon teachers and students to meet external expectations (Kelly, Meuwissen, Vansledright, 2007; LaBoskey, 2006; Phelps, 2006).

Considering that each local district had autonomy to make respective decisions on how to implement the state's reform (i.e., local control), understanding the processes to prepare teachers and students for adaptions toward alternative assessments was of utmost interest. The purpose for this study was to develop a descriptive account of one large Virginia school district's plan for implementation of alternative, locally developed assessments designed to supplant standardized measures. As policy reform with alternative assessments has been under-researched for the past 30 years, there is a need for studies conducted at the

district/state level that examine new methods and procedures to assess cognitive growth and complex skill sets (Darling-Hammond, et al., 2013). Although this study focuses on a single district, Longridge City Public Schools (pseudonym), I offer implications for diverse audiences on local, state and national levels, such as school division leaders, policymakers, and researchers, each with unique needs and a vested interest in alternative assessment as a reform initiative.

This research was timely as the action plan implemented by the Longridge City Public School division during the first trial, 2014 - 2015 school year, was reported. For this descriptive case study (Stake, 1995; Yin, 2014), the research questions are as follows:

1. What are the processes of a school division's leadership team for the development and enactment of alternative assessments?
2. What are the needs and challenges of a school division's leadership team during the development and enactment of alternative assessment?

The present study holds significance as it extends previous investigations of how state policy reform is linked with educational practice at the district level. In light of the reform, placing the focus on support and leadership at the local level was purposeful in discovering the considerations and actions necessary to meet the needs of numerous stakeholder groups across Longridge City Public Schools. To answer the research questions, the Virginia Department of Education 2014 legislative guidelines served as a framework to guide the development of this study. The tradition of case study (Yin, 2014) as an analytical approach was used to uncover the processes enacted by central office personnel (i.e., district leaders) to develop and implement alternative assessment, in the format of "performance-based assessments," or PBAs. Later, the findings reveal the district's ability to comply with the state's legislative reform mandates, while to navigate new territory of "local control." I now turn to an overview of the literature on the history of assessment in the United States.

LITERATURE REVIEW

As educational objectives are revised to reflect more complex thinking skills necessary for future success in college and career, alternative formats of assessment become more appealing (Basturk, 2005; Stecher, 2010). In this case, performance assessment formats become more suitable, as compared to traditional means of multiple choice, to assess today's 21st century learning goals (Conley, 2014). Within the past century, however, reform initiatives through curriculum, standards, and instructional methods have occurred at the national and state levels, reflecting varying ideologies on how to best prepare students for citizenship, the workforce, and higher education (Kliebard, 2004; Schiro, 2013).

Trends in State and Local Policy Reform

Although the topic of alternative assessment may appear contemporary as it relates to reductions in standardized testing (VDOE, 2014), such assessments in the late 1980s and early 1990s were widely used in state assessment systems for similar reasons during a shift toward competency of higher-level skill sets (Darling-Hammond et al., 2013; Kane & Mitchell, 1996). During this time, several states and districts across the nation consciously chose to incorporate aspects of higher-order thinking and processing skills into performance-based, alternative assessment systems. For instance, under the California Learning Assessment System (CLAS), 8th grade math students responded to two open-ended tasks with authentic situations followed by a multiple-choice section to assess mathematical cognition (Herman, Klein, Heath, & Wakai, 1994). In light of the new reform, a majority of teachers (91%) relied on the support through preparation materials provided by the state – the CLAS Mathematics Sampler – complete with sample problems teachers and students could use to become better acquainted with the assessments. As for writing, the California Writing Project served as a state and local district initiative to which training cadres of teachers with ‘analytic scoring schemes’ and effective models of writing focused on wide-scale writing across the curriculum (Herman, 1992).

The high stakes Kentucky Instructional Reform Information System (KIRIS) was

implemented to measure four main academic learning targets: (1) communication, (2) conceptual application, (3) real-world critical thinking and problem solving, and (4) and interdisciplinary learning (Stecher, Barron, Kaganoff, & Goodwin, 1998). KIRIS incorporated performance-based tasks with constructed responses across all core subjects in grades 4-5, 7-8, and 11, coupled with portfolios in mathematics and writing. Ample support from the state was provided at the local level through supplementary curriculum resources, scoring workshops using KIRIS assessments, and numerous opportunities for professional development (i.e., curriculum, instruction, and portfolio development).

Washington in 1998 adopted a 10-year longitudinal approach of alternative assessment reform starting with grades 4 and 7 as part of an assessment reform initiative known as the Washington Assessment of Student Learning (WASL) (Stecher, Baron, Chun, & Ross, 2000). A unique element in the preparation and implementation of the reform at the local level was the introduction of classroom-based assessments (CBAs), an effort to establish a system of balanced assessments in classroom practice. The state distributed assessment Took Kits complete with models of tasks, scoring guides, strategies and protocols for implementation, and even content frameworks to align lessons with WASL goals. Moreover, financial assistance was offered to schools that increased performance-based measures in practice (e.g., projects, experiments, oral presentations), regional learning centers were designated across the state for workshops, and research-based learning guides were supplied to all districts to disperse among schools.

States chose to adopt different approaches with alternative assessment based on circumstances and conditions. Firestone, Mayrowetz, and Fairman (1998) examined how state policies were locally developed and interpreted in Maine and Maryland, two states with moderate stakes attached to open-response, performance-based assessments. Maryland's School Performance Assessment Program (MSPAP) adopted a systematic, uniform approach with changes, whereas the Maine Educational Assessment (MEA) called for an individualized

approach at the local level. It was concluded that the MSPAP assessment items and scoring guides appeared to be more interdisciplinary and ‘open ended, allowing for greater application and invention’ (p. 108). A significant difference between states was the development of district-level training opportunities related to respective state assessments. Maine’s professional development activities were perceived as ‘causal events’ whereas Maryland districts used in-service days to provide collaborative time and resources for enhanced alignment between MSPAP and instruction.

Vermont’s decision to implement a systematic approach with portfolio-based measures was deemed unique and innovative at its launch during the early 1990s (Koretz, Stetcher, & Deibert, 1992). Teachers at the classroom-level compiled student performance evidence throughout the year for the subjects of mathematics and writing in grades 4 and 8. Volunteer teachers were trained at statewide meetings to score portfolios using criterion rubrics. During its early years, compiling and scoring portfolios were viewed as ‘burdensome’ with time, resources, money, and stress, despite the positive effects the reform had on teachers’ practice. Variations of complexity were found and there was ‘difficulty of training large numbers of raters to a sufficient level of accuracy’ (p. 12). However, Vermont’s actions demonstrated genuine interest regarding teacher practice at the ground level. Despite states’ initiatives and intense efforts with assessment reform, significant decline ensued from challenges with high costs, lack of educator support, time, scoring consistency, and criticism of certain elements of assessment implementation (Darling-Hammond & Adamson, 2010; Khattri et al., 1996).

Local Development and Enactment

Development of alternative assessment requires a collaborative effort among various stakeholders at the district level. For instance, Moon and colleagues (2005) conducted a 5-year study with 46 educational professionals (i.e., teachers, professors, curriculum coordinators, and department officials) to create alternative assessments in states using traditional, multiple-choice tests as their primary means of assessment. Central to their findings was a newfound

understanding among teachers and students of the criterion rubric as a learning tool. With new purpose, teachers viewed it as a resource to cross-check standards during initial planning, while students saw its usefulness to deepen their thinking and provided clear expectations.

Assessment reform also has implications for establishing new organizational routines and leadership opportunities at both the district and school levels. The efforts of Spillane, Parise, and Sherer (2011) within Chicago public schools add weight to the concept of 'coupling' policy with classroom practice. Their work demonstrated the significance in school-level collaboration between administration and curriculum coordinators – to develop intermittent assessments in grades 1 through 8 (i.e., mathematics, reading, and writing) – which spurred teachers' increased actions with data-informed decision making and creating student achievement benchmarks. To foster faculty leadership with curriculum and instruction, administration created instructional skills charts for teachers to track the alignment between their assessments, the standards, and the essential skill sets within their discipline. Such actions called upon teachers to collaborate with best practices during their regularly scheduled grade-level planning times. In both instances, newly established routines strengthened policy implementations within classroom practice and brought greater transparency to teaching and learning.

Portfolio-based approaches have historically served as a purposeful means of documenting student learning. At the request of the superintendent in a Long Island, New York school district, a multi-year project by Martin-Kniep, Sussman, and Metzer (1995) was conducted to explore the role of portfolio assessments. Significant to the positive changes in teachers' practice was the researchers' support through professional development workshops. The workshop space provided presentations on authentic assessments and provided facilitated support with crafting criterion rubrics to assess student learning. Afterwards, the teacher participants moved on to become district-wide trainers and impact teacher practice on a larger scale. Similar in nature, Kuh and Nelson (2014) in partnership with Making Learning Visible

(MLV), a project of Harvard Project Zero, were successful in creating a new culture of alternative assessment classroom practices which captured student learning through a portfolio-based approach. Significant to the outcome was the facilitated support through the MLV program which helped meet teachers' needs as they endured outside pressures of the Common Core State Standards, among other accreditation and program demands. In both studies aforementioned, the additional support provided salient contributions to participants' success.

Local District Support

With respect to policy reform, changes made to assessment formats are only part of the answer to improving teacher instruction and student learning (Herman, 1992). An assessment becomes most effective only when teachers are proficient with the knowledge and skills to interpret data and use results appropriately to modify instruction (Kelly et al., 2007; Perie, Marion, & Gong, 2009); therefore, oversight and support through professional development training are needed (Herman, 1992). Prior to making data-informed decisions on a larger-scale at the district level, professional learning interventions are necessary to build common understandings and beliefs about assessment data (Wayman, Jimerson, & Cho, 2012). Such professional development provisions should focus on enhancing teacher quality as an investment in student learning (Kuijpers, Houtveen, & Wubbels, 2010) and guidance in navigating the territory in which assessment, teaching, and learning converge (Dann. 2014). As in this case, research in the early stages of a new state or local assessment program was warranted (Stone & Lane, 2003). As such, this study seeks to add to the existing body of performance-based research, particularly as district leaders strive to develop and enact alternative measures and support the efforts to implement the 2014 Virginia legislative policies within practice.

METHOD

The tradition of case study in this qualitative research study was grounded in the expansion of educational theory (Yin, 2014). For the purpose of this study, one of the largest

Virginia school districts, Longridge City Public Schools – with over 70,000 students attending more than 80 elementary, middle, and high schools in the school district – served as the unit of analysis. Specifically, the district's Department of Teaching and Learning located in the district's central office served as *the case*, operating as a bounded system comprised of select district leaders striving to develop understanding of the legislation's impact at the local level (Creswell, 2013). The events specific to the boundaries, individuals, and processes of the district as the single case were carefully documented (Yin, 2014).

Participants and Procedures

An initial examination of the division's responsive plan identified the main division leaders, those of whom were (1) actively involved in overseeing the implementation of the locally developed alternative assessments, and/or (2) managing the development processes of the assessments for either elementary grades, secondary grades, or both. Out of the five central office personnel leading the charge for the division, four were available to serve as participants. A cover letter, informed consent, and demographic sheet were sent to each participant in advance that clearly outlined the intentions of the study.

The selected participants were specifically selected as their central office positions supervised the five academic areas, at the elementary and secondary levels, impacted by the legislation. The Executive Director of Secondary Teaching and Learning, Dr. Lane (pseudonym), managed curriculum in grades 6-12, among several other division responsibilities pertaining to technology and the arts. The division's Chief Academic Officer, Dr. Jonas (pseudonym), directed The Department of Teaching and Learning with the assistance of four Executive Directors and three Directors, who directly oversee their respective offices within the Department of Teaching and Learning. The Secondary Social Studies Coordinator, Ms. Wright (pseudonym), improved instructional practices, to include writing curriculum, and is tasked with establishing and facilitating various committees. Finally, the Executive Director of Elementary Teaching and Learning, Dr. Cahn (pseudonym), managed curriculum in grades K-5, among

several other division responsibilities pertaining to Title I and II. Although participants were relatively new to their roles, each had served in the field of education for over 20 years.

Data Collection

To uncover the effects of the state legislation within the district, data were first collected from individual, semi-structured interviews lasting approximately 40 minutes in length. Interview protocol questions were sequenced in a manner that provided rich, descriptive coverage of the two research questions in this case study (see Appendix A). Examples of questions included, “If I were to look at a timeline for the steps involved in this trial school year of 2014-2015, what might it look like?”, “Can you describe any challenges in the district’s plan to administer alternative assessments?” and “How will results of the alternative assessments be communicated?” During the interviews, probing questions such as, “What was the importance of …?” and “Can you tell me more about …?” were used to establish clarity and seek further elaboration for deeper understanding.

By converging findings from multiple data sources for triangulation purposes, I aimed to increase confirmation and credibility (Stake, 2005; Yin, 2014). Field notes in this case were interpretive, bringing chronological and sensory meaning to all personal accounts (Bogdan & Biklen, 2003) and were used to exhibit depth and detail of the setting (Patton, 2002). Context from field experiences are briefly described below for the purpose of bringing greater meaning to supplement the findings. Additionally, key documents that surfaced during the investigation were examined mainly for empirical knowledge to create deeper understanding of the case (Bowen, 2009).

Field observation: Regional meeting. There is a history of communication and collaboration among the 12+ districts that make up their respective region in Virginia. In early 2015, a \$20,000+ grant was awarded from the state, allowing the region to solicit professional development support from region universities. Field notes were taken during a March 2015 Virginia regional session, to which district-level coordinators from the three disciplines (i.e.,

social studies, science, English/Language Arts) and teachers attended professional development on assessment literacy.

Field observation: Local school board. The Longridge City Public School district's school board, superintendent, and central office personnel gathered monthly to conduct briefing and training sessions. The relationship between the state's mandates outlined in the 2014 legislation division's plan and the district's action plan served as the highlight of the May 2015 workshop. Field notes were taken while observing the session devoted toward the discussion of performance-based alternative assessments (Bogdan & Biklen, 2003). Mainly, the purpose of this meeting was to share, analyze, and discuss the locally-developed alternative assessments designed to supplant the tests removed.

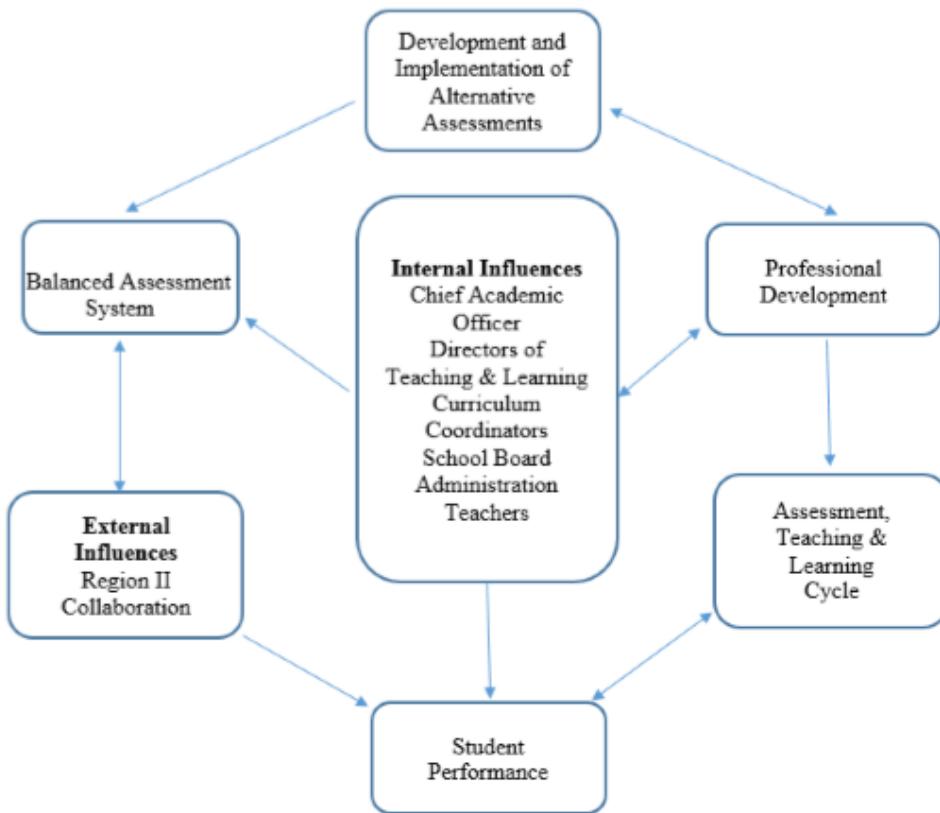
Artifacts: Training documents. Common among the collected documents was a purpose or intent to educate the audience on the legislative mandates, the design, development, and implementation processes of performance-based assessments. Among the collection were: resource materials featured at the region training workshop, a presentation from the Regional Instructional Leaders Meeting (property of the Virginia Department of Education), and a presentation on historical thinking skills used to train 6th/7th grade task development committee members at the district level.

DATA ANALYSIS

Categorical aggregation, direct interpretation, and pattern identification were used to develop and define broad categories emerging across interviews (Stake, 1995). In the first stage, relevant data to the research questions were categorized, to which interpretations through participants' direct quotes were used to present greater context. In the second stage, patterns, identified through common traits among the categories were collapsed into defined themes, accompanied by the most relevant participant quotes provided as supporting evidence. Causal links, or rationales for the decisions made and actions taken to meet the state legislative mandates, were captured through descriptive analysis (Yin, 2014). To provide a visual

representation of the relationships among the findings, a causal network display is featured (see Figure 1).

Figure 1. Causal Network Case Display (Miles & Huberman, 1994)



To obtain rigor and credibility within the analysis process, a peer debriefer was solicited for trustworthiness purposes. The collaborative dialogue between myself and the debriefer was valued for the opportunity to reflect in a manner that questioned the role of alternative assessment in the study, and to better understand how influences of the researcher's experiences could strengthen credibility (Lincoln & Guba, 1985). Selection of the debriefer was based on her numerous years of qualitative research experience and her extensive practice with portfolio-based alternative assessments. As an outsider to the case, she brought an unbiased perspective into the debriefing process. During the collaborative sessions, questions such as

“How does this relate to...?” and “Why might this be important?” aided in the focus of attaining the goals outlined by the research questions (Spillett, 2003).

Additional strategies of trustworthiness were used to maximize validity, with emphasis on establishing credibility, authenticity, and substantive validation (Lincoln & Guba, 1985). For instance, member checking was conducted by asking participants to validate the researcher’s interpretations in the transcripts (Johnson, 1997; Mays & Pope, 2000). Responses of acceptance were received from all participants, along with a single request for minor refinements of Ms. Wright’s transcript. During interviews, negative case analysis was used to speculate and gain greater insight of experiences with other districts for comparison (Patton, 2002).

FINDINGS

The descriptive nature of the case (i.e., the district) was bounded by time, place, and circumstances within the trial school year of 2014-2015 (Yin, 2014). Findings illustrated the school district’s experiences in navigating new territory to adopt policy reform into educational practice. New understanding and meaning derived from the collaborative encounters of division participants (e.g., department leaders, curriculum coordinators, teachers, and School Board members) and were socially-constructed through collected words, testimonials, and documented actions. The cohesive relationships that developed between the research questions and emerging themes are addressed through generalizations of what was learned (Creswell, 2013). Each data collection phase brought new findings, providing a deeper understanding of the division’s timeline of events, considerations, and decisions related to the state’s mandates. Three primary themes are presented in a chronological, descriptive fashion, complete with participants’ voices as “abbreviated vignettes” to highlight significant explanations of key case study issues (Yin, 2014).

Establishing a Division-Wide Plan of Action

The district's initial response to the legislation called for the establishment of a road map to analyze where the division has been, was currently, and planned to advance. Motives to think and act in a specified way derived from the division's long-term Strategic Plan, revealing (1) the state of readiness in relation to curriculum and assessment, as well as (2) the experienced benefits and challenges during the development and implementation phases of the alternative assessments.

The historical path. Prior to the legislation, Longridge City Public School's strategic plan for student success established educational goals and priorities planned around "balanced assessments," particularly the building of authentic performance tasks into curriculum. Specifically, the Departments of Teaching and Learning "began creating a K-12 matrix of performance assessments to outline where they had embedded [performance tasks] in the curriculum, and in which content areas" (Dr. Jonas). Starting in the 2013-2014 school year, and continuing into 2015-2016, Dr. Jonas explained the division's plan:

The support will be to help teachers understand the power of a balanced assessment system and what it truly means, what it looks like and feels like in the classroom, and not just as it relates to the SOL test changes but just in general as a best practice.

In support of the expansion with best practices, Dr. Lane stated, "We've been talking about this for a long time and this just gives us the impetus to say, Ok, we don't have a choice now, this is where we have to go with it." To support the district and state visions, Dr. Cahn explained "When you look at the state's recommended replacements, one of the alternative assessment options was performance-based assessment, and that's where we were all going with our curriculum anyway."

The division's present awareness. Providing assessment evidence for each eliminated SOL test became the division's priority, as students now self-constructed their responses to demonstrate understanding, in contrast to selecting a letter. Acknowledging this

new reality, Dr. Lane commented, “from the moment the [State] Superintendent’s memo was released to school districts, stating that the legislation was now officially in action, the division needed to develop a plan of action that was in accordance with the state’s mandates.”

Speaking on behalf of the Department of Teaching and Learning, Dr. Cahn stated “it was imperative for us to not only know what was available and know what we had to create, but then also be able to communicate this with schools in a timely fashion.” According to Dr. Jonas, there was significance in helping to shape the understanding in a large school division,’ particularly in response to what this will look like for schools.

To begin, both elementary and secondary supervisors conducted an internal audit which became a threefold process: 1) examine the current curriculum to see what types of assessments already existed within the curriculum, 2) recognize which of those assessments matched the state’s recommended guidelines for the alternative assessment, and 3) determine where the needs lie. For instance, when elementary curriculum is taught with fidelity, the standards-based report card in grades 3-5 served as just one record that the content standards were taught and assessed. Similar for grades 6 and 7 social studies, open-response quarterly assessments have traditionally been used by teachers to serve as just one record of content strand coverage. The aim was to move beyond these measures through additional, formative assessment approaches.

District Initiatives Based on the Legislation

Specific to meeting the legislative mandates was the development and enactment of the alternative assessments with teachers in practice. To that end, initiatives were taken by the district leaders to best prepare teachers for the changes that ensued in transitioning to a new assessment system. Throughout the 2014-2015 school year, professional development was a key component in this process, and occurred at both the district and region levels. Workshops and trainings were designed for select attendees with varying intentions. The agendas were similar at both levels, taking on a social-constructivist approach to establish meaning of

alterative assessment and find common ground during the construction of the performance-based tasks. Strategic efforts were made by the district leaders to involve teachers and neighboring districts during this development.

Teacher collaboration: District level. To begin, on January 18, 2015 a memo was released to schools inviting 6th/7th grade social studies teachers to apply and become an Alternative Assessment Development Committee member. Responsibilities included assisting in the development and refinement of alternative assessments (i.e., performance tasks) and accompanying rubrics. ‘We always make sure we involve teachers when building the assessments’ (Dr. Lane). Additionally, Ms. Wright added ‘we’re relying on teachers to come along and assist with the writing of this...some of the ones we had in place, teachers helped tweak.’ To support further growth in this process, one 6th and one 7th grade teacher from the development committee attended a professional development conference at the regional level.

During a school board training session, the directors presented samples of the locally-developed alternative assessments in all five areas to the superintendent and school board members. Words such as “real-world,” “higher-level skills and processes,” and “analyze and interpret” were used during the discussion of the tasks. School board members in particular remarked on the completion of the tasks and how it would be necessary for students to use writing, research, and communication skill sets, make inferences, and draw conclusions, to be successful in crafting their responses. During the discussion of changes in teachers’ practice, Ms. Wright stated how social studies teachers, for instance, had been working diligently with literacy components in social studies, particularly in historical writing.

District collaboration: Region level. Starting in February of 2015, collaboration began among Virginia’s regional districts to support one another in not only meeting the state mandates outlined in the legislation, but more importantly, to provide professional support to the stakeholders directly involved in the enactment. At a Virginia Regional workshop session, curriculum coordinators represented the 12+ districts, to which each brought four teacher

representatives from the elementary and secondary subject areas impacted by the legislation. Significant to this training was the leadership from a professor at a well-renowned, nearby university. Throughout a full day of professional development on the topic of assessment literacy, his expertise in assessment was used to unpack the meaning and understanding of alternative-based, formative assessment. There was ample time for members from various school districts to collaborate and generate ideas for alternative assessments, respective to the needs of each district's staff and student needs.

Differentiation. A conscious effort was made to recognize what was in the best interest of teachers while making the adaption. Both elementary and secondary Departments of Teaching and Learning considered differentiation for teachers with regard to (1) teachers' readiness with pacing, and (2) teacher options or 'choice' of assessments to implement with their students.

Aligning the vision of the Department of Teaching and Learning with the thinking of teachers was a chief consideration. For example, should teachers sway from the recommended assessments in the curriculum, "the challenge becomes, why?...Is it because you have other ones you've used for years that are meeting your same needs?" (Dr. Cahn). Additionally, Ms. Wright shared how "some teachers were concerned about the time that it takes to deliver these types of assessments and the time it takes to grade them." While several participants commented on teachers' excitement about the assessments, there was also recognition of teachers' additional classroom challenges, such as student levels of literacy. In response, the shaping of future implementation of the alternative assessments responded to the support teachers needed at various levels of readiness. For instance, with respect to fifth grade writing the district invested in *Being a Writer*, a scripted writing program provided as a supportive resource for teachers. It was explained, "You can teach the curriculum of fidelity and use the assessments that are already embedded into the curriculum, or you can use this one (*Being a Writer*) too, or you can use all of them" (Dr. Cahn). Acknowledging the degree of flexibility at the

secondary level for sixth and seventh grade social studies, Dr. Lane stated the continued plan was to “develop multiple assessments for each of the strands so that a teacher has a choice of which type they want to use, but all are performance-based assessment.”

Local communication and collaboration. Clear and consistent communication served as the linchpin in educating various district stakeholders impacted by the assessment reform. The district leaders acknowledged their efforts to effectively communicate the details of their action plan, first through a Principals’ Memo sent out to administrators across the district, and then to teachers through multiple rounds of emails. According to Dr. Lane, “It was previously talked about at league meetings with middle school principals, so they knew it was coming.” Furthermore, Ms. Wright shared her effort as the social studies coordinator to “put them (performance tasks and rubrics) in the [district’s database]… and asked teachers to print them and deliver them as they [the performance tasks] came up in the curriculum.” To aid in teachers’ responsibilities, Ms. Wright collaborated with two content leads from each building at least once per quarter. These collaborative meetings were not only significant to the district-wide communication, but they provided a platform for teachers to field their questions and receive answers during the transitional phase in the spring of 2015.

Implementing Balanced Assessment: The Cycle of Teaching and Learning

Bound within the division’s plan was a vision for alternative assessment to become a valued component within the cycle of teaching and learning. The beliefs and values associated with alternative assessments as a best practice influenced the specific goals, or the standards of success, laid out by the district. Furthermore, conscious efforts were made to ensure the alternative assessments were measured and reported through a systematic approach.

Thinking skills and processes. Disciplinary literacy was a critical element of consideration during the development of alternative assessments. For example, the work of the Stanford History Education Group (SHEG) was frequently noted during interviews and observations for authentic, historically-based lessons with historical thinking skills embedded.

With such tasks serving as models for social studies, Dr. Lane stated how the “focus is on breaking away from fact-driven social studies curriculum and instruction...we knew this would be a great fit for where we are going.” Specifically, Dr. Lane and Ms. Wright referenced the historical thinking skills of sourcing, corroboration, and contextualization embedded in the locally-developed social studies performance tasks. The tasks were modeled after those featured from the SHEG, designed to curb memorization of facts and encourage analysis and historical thinking through a more critical lens. When students engaged in the newly designed tasks, Ms. Wright explained the gains for student thinking:

They’re able to open up a browser and search for information and be able to say, Is this information good? And because I have the skills to source this information, because I can question who wrote it, ‘What was its purpose? Are there other pieces of information elsewhere that corroborate this information?’

The locally developed alternative assessments. Common among the elementary and secondary alternative assessments were elements such as decision-making, problem-solving, communication, and literacy. Each required open-ended, constructed responses, and was accompanied with a criterion rubric offered to students and teachers alike. The nature of the tasks was described as “authentic,” as they included primary source documents and “real-life subject matter,” allowing students to make relevant connections between the past and present.

At the elementary level, one particular third grade social studies task spanned the entire school year. For instance, students began by brainstorming problems in the classroom or community, and then hypothesized possible solutions. The teachers worked with students on finding real world solutions that could solve the problem, and finally create a product. The value in this type of alternative task was described by Dr. Cahn:

It’s a whole year of the citizenship and community...This is what you do when you have a problem...This is how you solve it’ so it’s a very ongoing real-world task. And by the

time the students get to the end of the year the students have solved something that's true to them.

Secondary social studies tasks, on the other hand, had an average of four conceptually-based, primary source document-based questions. In terms of the teacher practice required in preparation for such assessments, Dr. Jonas noted how "students are benefiting from the fact that they're probably getting instruction that feels a little more authentic and less like drill-and-kill," assuming teachers' were making adjustments in their practice to mirror the types of questions students were asked to craft responses for. Additionally, the secondary social studies performance tasks required student analysis, such as having them analyze historical artwork indicative of socio-political reform. In comparison with the SOL tests, Ms. Wright described the grade 6 and 7 performance tasks as "certainly more analytical ...more performance-based and authentic for their (student) learning." Interdisciplinary writing in particular was a tenet of the district's action plan, and would continue to receive greater attention at both the elementary and secondary levels.

Frequently expressed among interviews was the desire for alternative assessments to serve as a catalyst to positively impact and improve teaching and learning. For example, Dr. Cahn noted, "You don't just use an assessment for a one shot deal, assessment should drive your instruction." The district leaders remained hopeful that increased use of alternative assessment, outside of what was mandated through the reform, would see its way into classrooms as teachers continued to collect formative data on their students' learning.

Collecting and analyzing student data. At the classroom level, teachers' ability to unpack curriculum and use various types of assessments to obtain meaningful student data, was a critical concern. Dr. Cahn shared how "it's really the whole learning to trust the assessment cycle...and when does a formative become a summative, and when do you reassess?" With respect to improving teacher practice, Dr. Jonas shared how "our teachers are benefiting from being able to move away from that high-stakes, data-driven type of instruction,

to data-driven to inform teaching.”

At the district level, in order to monitor and gauge the progression of the district-wide assessments, a plan was enacted through the Department of Teaching and Learning to collect a sampling of scored tasks from each school. According to Dr. Lane, “What we’ve asked them to do is to give us a high, a medium, and a low, so we can take those and look at interrater reliability with how they (teachers) are grading on the rubric.” At the secondary level (grades 6 and 7) “that would be nine from each grade level...and that’s 18 total for each school” (Ms. Wright). There was an expressed desire to have models for each of the levels for use during professional development training offered throughout the summer.

Future considerations. The trial school year of 2014-2015 served as a reflective process for the division. For instance, Ms. Wright revealed considerations for next steps in the refinement process for the following school year:

There's room for improvement, but I think it's a good year to do this and know that it's not going to be perfect, but it's also a good year to say, 'What are we still missing? What needs to be addressed? What kinds of professional development needs to happen so that our students can perform and achieve better?'

A state of mindfulness was evident upon considering the future refinements of the district's action plan. For instance, possibilities of how teachers would perceive the steps embedded within the district's plan were carefully considered by the district leaders in advance through various scenarios and solutions. Furthermore, the needs for future interventions to make impactful changes in teachers' practice in support of the reform were also a concern, to which professional development would continue to play a significant role.

DISCUSSION

Of significant interest to this study's findings was the variable of “local control” with the state's assessment policy, and how each district managed this new authority. In this case, the actions of central office personnel (i.e., district leaders) to conduct a full analysis of where

performance-based assessments were present in the existing curriculum greatly supported the legislative reform guidelines (VDOE, 2014). This analysis was not limited to the areas in which an SOL test was removed, and instead included a widespread cross-check in the K-12 continuum for student exposure to and practice with alternative, performance-based assessments. In other words, the district's action plan had implications for enhancing student growth and teacher practice on a larger scale and to a greater capacity. Notably, the reality of implementing alternative assessments served as just one piece of a larger picture – the alignment between the policy reform and the division's philosophy of essential relationships between assessment, curriculum, and instruction, consciously bound within a cycle of teaching and learning (Dann, 2014; Perie et al., 2009; Solley, 2007; Stecher et al. 2000).

The vision for alternative assessment implementation was clear in terms of where teachers and students needed to be, when, and how they were going to get there. First, a conscious effort was made to engage teachers with best practices of data-informed decision-making as evidence by their submission of scored student samples (Spillane et al., 2011). Secondly, central office personnel demonstrated a robust understanding of the district's readiness for the alternative assessments by initiating a plan to include a teacher committee and regional collaboration in the creation of performance-based tasks, both of which called for a certain degree of assessment literacy (Khattri et al., 1998). For instance, authentic, real-world contexts were valued as essential elements incorporated into the tasks, such as with service learning (McDonald, 2008). Additionally, criterion rubrics were designed as effective teaching and learning tools (Moon et al., 2005), and interdisciplinary opportunities – such as with history, writing, and the arts – were integrated to capitalize on learning experiences (Firestone et al., 1998; Stecher et al., 1998; VDOE, 2014).

The present and future professional needs of teachers and their practice were carefully considered throughout the process of enacting the state's mandates. In making the transition from the old to a new assessment accountability system, provisions for support in assessment

design, implementation, and scoring were strategically planned for in the summer's professional development workshops (Kuh & Nelson, 2014; Martin-Kniep et al., 1995). Such practice and opportunities to score student tasks through data-driven processes serve as an investment in teachers' professional growth and aid in their ability to make informed pedagogical decisions (Wayman et al., 2012). Overall, forward thinking on behalf of the district leaders was evidenced through the district's Strategic Plan to recognize alternative assessments within a system of balanced assessments, bound within the teaching and learning cycle (VDOE, 2014).

Limitations

The effects of the 2014 legislation were circumstantial for each local district in the state of Virginia. The findings of this study were bound within the district as a result of the decisions made by central office personnel that affected a myriad of stakeholders throughout the district (Yin, 2014). There were few potential interviewees in central office responsible for developing and enacting the division's action plan, limiting the number of participants' voices. To overcome such constraints, field observations were conducted; however, minimal opportunities arose during the study's window during the later months of the first trial, 2014-2015 school year. Lastly, a significant challenge in this study included the lack of time and collected evidence to suggest if alternative methods could serve as an effective means to assess student proficiency of content knowledge and skill sets, particularly when compared to the traditional, standardized approach.

Future Research

This study addressed how alternative assessments can be woven into curriculum in grades 3-7; however, descriptions of everyday classroom practices were not present. Observations of teachers' practice in the classroom, analysis of lesson plans, and/or focus group teacher interviews should be considered in future research to gain deeper understandings of how ongoing data are used to make instructional decisions. Strategies, interventions, and

daily tasks that reinforce ongoing practice with the skills and processes featured in alternative assessments should be closely analyzed.

Researchers interested in the relationships among this study's findings, in relation to theory and research, can bring greater attention to this topic, particularly when addressing the missing elements aforementioned that could provide fruitful directions along a similar research path. Much can be learned from leading states during the performance-based assessment movement in the late 1980s and early 1990s (Darling-Hammond & Adamson, 2010; Darling-Hammond et al., 2013; Khattri et al., 1996). Only time will tell if additional states will choose to entertain opportunities with alternative assessment reform and thrive as a result from the lessons of our past.

Conclusions

This descriptive account on an important, trending topic has potential to serve as a stimulus among various targeted audiences to draw much needed attention to phenomenon linking policy reform with educational practice. Contemporary findings from this study provide a deeper understanding of the 'why' and 'how' behind one district's initiatives to adopt alternative assessments designed to supplant select standardized tests (Yin, 2014). Naturally, there are implications for the district in this study, particularly as central office personnel continue to monitor practices aligned with the state's mandates and use findings to inform future actions. However, the findings can certainly speak to other potential audiences with a vested interest in alternative assessment as a reform initiative.

Upon introducing educational policy into practice, local district leaders can look critically at the actions, considerations, cautions, and reflections made by the Longridge City Public School district while establishing a responsive action plan. This study highlights the importance of understanding the state of readiness of all stakeholders involved in the planning and implementation phases. Communication and collaboration among central office personnel,

curriculum coordinators, department heads, local school board, and administrators are essential to a division's success.

Findings should be shared with policymakers at-large as a communication device to host conversations about possibilities with assessment reform (Yin, 2014). In consideration of adopting alternative assessments on a larger scale, leaders may first benefit from becoming familiar with the protocols and procedures outlined in the 2014 Virginia legislation. External support among division schools can be impactful on the professional growth of teachers, and ultimately student learning. As alternative assessment is not a new idea, state leaders can look critically at lessons from the past to attempt implementation in K-12 settings and bring forth positive change in assessment practices.

REFERENCES

- Bogdan, R. C., & Biklen, S. K. (2003). *Qualitative research for education: An introduction to theories and methods* (4th ed.). Boston, MA: Allyn & Bacon.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40.
- Conley, D. T. (2014). Common core development and substance. *Social Policy Report*, 28(2), 1-15.
- Creswell, J.W. (2013). *Qualitative inquiry and research design: Choosing among five traditions* (2nd ed.). Thousand Oaks, CA: Sage.
- Dann, R. (2014). Assessment as learning: blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149-166.
- Darling-Hammond, L. & Adamson, F. (2010). *Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning*. Stanford: CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Herman, J., Pellegrino, J., et al. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Firestone, W. A., Mayrowetz, D., & Fairman, J. (1998). Performance-based assessment and instructional change: The effects of testing in Maine and Maryland. *Educational Evaluation and Policy Analysis*, 20(2), p. 95-113.

Herman, J. L. (1992). *Accountability and alternative assessment: Research and development issues*. CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

Herman, J. L., Klein, D. C., Heath, T. M., & Wakai, S. T. (1994). *A first look: Are claims for alternative assessment holding up?* CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

Johnson, R. B. (1997). Examining the validity structure of qualitative research. *Education*, 118(2), 282-292.

Kane, M. B., & Mitchell, R. (1996). *Implementing performance assessment: Promises, problems and challenges*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kelly, T., Meuwissen, K., & Vansledright, B. (2007). What of history? Historical knowledge within a system of standards and accountability. *International Journal of Social Education*, 22(1), 115-145.

Khattri, N., Reeve, A. L., & Kane, M. B. (1998). *Principles and practices of performance assessment*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kliebard, H.M. (2004). Curriculum ferment in the 1890s. In *The struggle for the American curriculum: 1893-1958* (3rd ed.). (pp. 1-25). New York: Routledge.

Koretz, D., Stetcher, B., & Deibert, E. (1992). *The Vermont portfolio program: Interim report on implementation and impact, 1991-92 school year*. Santa Monica, CA: The RAND Corporation. Kuh, L., & Nelson, C. (2014). Navigating policy and trying out new ways of teaching: Teachers and children becoming experts. *New Educator*, 10(1), 62-69.

Kuijpers, J., Houtveen, A., & Wubbels, T. (2010). An integrated professional development model for effective teaching. *Teaching & Teacher Education*, 26(8), 1687-1694.

LaBoskey, V. K. (2006). "Reality check": Teachers' lives as policy critique. *Teachers and Teaching: Theory and Practice*, 12(2), 111-122.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Martin-Kniep, G. O., Sussman, E. S. & Meltzer, E. (1995). The north shore collaborative inquiry project: A reflective study of assessment and learning. *Journal of Staff Development*, 16(4), 46-51.

Mays, N., & Pope, C. (2000). Qualitative research in health care: Assessing quality in qualitative research. *British Medical Journal*, 320, 50-52.

McDonald, B. (2008). Assessment for learning in project based learning. *International Journal of Learning*, 14(10), 15-27.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: Sage.

Moon, T., Brighton, C., Callahan, C., & Robinson A. (2005). Development of authentic assessments for the middle school classroom. *Journal of Secondary Gifted Education*, 16(2/3), 119-133.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.

Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5-13.

Phelps, R. P. (2006). Characteristics of an effective student testing system. *Educational Horizons*, 85(1), 19-29.

Schiro, M.S. (2013). *Curriculum theory* (2nd ed.). Thousand Oaks, CA: Sage.

Solley, B. A. (2007). On standardized testing: An ACEI position paper. *Childhood Education*, 84(1), 31-37.

Spillane, J. P., Parise, L. M., & Sherer J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, 48(13), p. 586-619.

Spillett, M. A. 2003). Peer debriefing: Who, what, when, why, how. *Academic Exchange Quarterly*. (7)3. Retrieved on March, 18, 2015, from
<http://www.thefreelibrary.com/Peer+debriefing%3a+who%2c+what%2c+when%2c+why%2c+how.-a0111848817>

Stake, R. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stecher, B. (2010). *Performance assessment in an era of standards-based educational accountability*. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Stecher, B., Baron, S., Chun, T., & Ross, K. (2000). *The effects of the Washington state education reform on schools and classroom*. CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

Stecher, B. M., Barron, S., Kaganoff, T., & Goodwin, J. (1998). *The effects of standards-based assessment on classroom practices: Results of the 1996-97 RAND survey of Kentucky teachers of mathematics and writing*. CSE Technical Report. Los Angeles: UCLA National Center for Research on Evaluation, Standards, and Student Testing.

Stone, C. A., & Lane, S. (2003). Consequences of a state accountability program: Examining relationships between school performance gains and teacher, student, and school variables. *Applied Measurement In Education*, 16(1), 1-26.

Virginia Department of Education (2014). *Guidelines for Local Alternative Assessments for 2014-2015*. Standards of Learning (SOL) & Testing, VA: Virginia Department of Education. Retrieved from http://www.doe.virginia.gov/testing/local_assessments/index.shtml

Wayman, J., Jimerson, J., & Cho, V. (2012). Organizational considerations in establishing the data-informed district. *School Effectiveness & School Improvement*, 23(2), 159-178.

Yin, R. K. (2014). *Case study research: Design and methods*, (5th ed.). Thousand Oaks, CA: Sage.

APPENDIX A: INTERVIEW PROTOCOL

Individual, Semi-Structured Interview Questions (10-12)

1. Can you describe any major factors that were taken into consideration in the beginning of this process?
2. If I were to look at a timeline for the steps involved in this trial school year of 2014-2015, what might it look like?
3. What are the roles and responsibilities of the stakeholders involved in the process of administering alternative assessments?
4. What collaboration, if any, is taking place among regional and/or state school districts?
5. Can you describe any successes in the district's plan to administer alternative assessments?
6. Can you describe any challenges in the district's plan to administer alternative assessments?
7. Why were performance tasks chosen as the format to assess students' content knowledge in social studies?
8. How are the authentic assessments designed to measure if students are making academic progress in social studies?
9. How will the alternative assessments be scored?
10. How will results of the alternative assessments be communicated (with state, central office, administrators, parents, students)?
11. How are assessment data intended to be used, if at all, by teachers?
12. What supports, if any, have been provided through professional development to assist in this process?

Probing Questions:

Negative Case Analysis-

In your opinion, what might prevent _____?

General-

How is the importance of ____ related to ____?

Why might ____ be important in the situation of ____?

Can you tell me more about ____?

Why might you suppose ____?